Detection of Geometric Structures by an Optimally Subsampled Shearlet System in Noisy Digital Images

Philipp Petersen

Abstract

We provide a statistical analysis of the ability of digitized continuous shearlet systems to detect objects embedded in white noise. We analyze the possibility to subsample the shearlet system and obtain a system of significantly reduced cardinality that can still yield statistically optimal detection results.

1   Introduction

Detection and classification of singularities in images is an important task in imaging applications. It has been shown in [3, 4] that the continuous shearlet transform can be applied to identify location and position of jump singularities of an image. These works have been extended and improved in [9] to include compactly supported shearlets, which have optimal spatial localization and can therefore handle singularities better.

All of the above results use a continuous shearlet system i.e. a system constructed from a generator $\psi \in L^2(\mathbb{R}^2)$ together with elements

$$\psi_{a,s,t}(x) = \psi(A_a^{-1} S_s^{-1}(x - t)), \quad a \in (0,1], \ s \in [-1,1], \ t \in \mathbb{R}^2, \tag{1}$$

where

$$A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix} \text{ and } S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}.$$

For applications we need to sample the parameters a, s, t and also restrict the functions $\psi_{a,s,t}$ to a digital lattice or an $n \times n$ pixel image.

In this work we will examine this digitization from a detection point of view, in the sense, that we aim to describe a suitable digital shearlet system to optimally detect geometrical structures embedded in noise. This is inspired by the work of Arias-Castro, Donoho and Huo [2], which describes how to construct statistical tests to test for some geometric structures such as lines, intervals or rectangles. There it is assumed that ξ is a discrete signal of length $n \in \mathbb{N}$ with certain characteristics such as jumps and a white noise signal $z(i)$, where i is an element of some discrete index set I_n and $z(i) \sim \mathcal{N}(0, 1)$. In the one dimensional situation we choose $I_n := \{1, \ldots, n\}$. Similarly in higher dimensions we let $I_n := \{i = (i_1, i_2, \ldots, i_k) : 1 \leq i_l \leq n, l = 1, \ldots, k\}$. The measured signal is now given by

$$x(i) = \xi(i) + z(i), \quad i \in I_n. \tag{2}$$

In [2] this model was used to detect discrete lines, rectangles, circles and intervals embedded in white noise. We will continue on this path and examine the optimal detection of structures

connected to discontinuities and continuous shearlet systems. Our main result, Theorem 3.6, states, that careful subsampling of a continuous shearlet system does not influence its abilities to detect certain geometric properties. This subsampling yields a reduction of the size of the system from $O(n^4)$ to $O(n^2 \log(n))$ for an $n \times n$ pixel image. The manuscript is structured as follows. We will describe the methodology of using statistical tests for the detection of geometrical structures in Subsection 2.1. Then we will proceed to construct a test to find jump sdiscontinuities in 1d signals in Subsection 2.2. From these methods we will extract a general framework in Subsection 3.1. This will then be applied to the continuous shearlet system in Subsection 3.2.

2 Detection of Jump Singularities

2.1 Problem Setup

We will begin with the detection of jump discontinuities. From the point of view of detection this geometrical structure is closely related to that of intervals and therefore we will first briefly recall the detection of intervals. Using the model (2) with $\xi = A_n \xi_{[a,b]}(i)$, where $\xi_{[a,b]}(i) = \sqrt{b-a}\chi_{[a,b]}$, $I_n = 1, \ldots n$, the authors of [2] showed, that a test can reliably the presence of an interval in an asymptotically stable way, that will be introduced below, from the signal

$$x(i) = A_n \xi_{[a,b]}(i) + z(i), \quad 1 \leq i \leq n, \tag{3}$$

if A_n lies above a critical threshold. On the other hand if A_n lies below this threshold, no test can reliably detect the interval. The phrase of asymptotically reliable detection needs some mathematical rigor and hence, we repeat the definitions of *asymptotically powerful* and *asymptotically powerless* tests.

Definition 2.1. *[2] In a sequence of testing problems* $(H_{0,n})$ *vs.* $(H_{1,n})$*, we say that a sequence of tests* (T_n) *is* asymptotically powerful *if*

$$P_{H_{0,n}} \{T_n \text{ rejects } H_0\} + P_{H_{1,n}} \{T_n \text{ accepts } H_0\} \to 0, \quad n \to \infty$$

and the sequence is asymptotically powerless *if*

$$P_{H_{0,n}} \{T_n \text{ rejects } H_0\} + P_{H_{1,n}} \{T_n \text{ accepts } H_0\} \to 1, \quad n \to \infty.$$

2.2 Detecting 1D Jump Singularities

We aim to construct a series of tests that are asymptotically powerful, to detect jump singularities in noisy data. This means, that we obtain data of the form

$$x(i) = \xi_J(i) + z(i), \quad 1 \leq i \leq n, \tag{4}$$

where ξ_J exhibits some sort of jump, and $z(i) \sim \mathcal{N}(0,1)$. In the spirit of [2] one must first pose the question under which conditions any test can reliably detect the presence of a jump in the signal.

First of all we need to determine what a jump singularity should be. At this point we may want to consider the prototype of a jump singularity, given by objects of the form $\psi(i) =$

$\chi_{[1,s]}(i), i \in I_n$ which model a jump at the position s from 1 to 0. These types of jumps are not very challenging, in fact the detection of these is already covered by the possibility to detect intervals as in (3). Another approach could be to model jumps simply as a point where the function value differs strongly from that of the neighboring points. Taking pairwise differences leads towards the problem of finding one elevated mean within white noise, i.e. *the needle in a haystack problem*, [6, 7]. In order to be reliably detectable this would require the jumps to be quite high and in fact higher than we might want to. In fact, a jump could also be detectable, if it is small but in a neighborhood apart from the jump the function values of the signal change very little. In other words, if before and after the jump we have an almost constant signal, then we would also expect to be able to detect smaller jumps. Indeed, a reasonable model for a jump singularity is that a function has a large Haar wavelet coefficient. This will now be our model.

Let us examine the continuous Haar wavelet transform on $1 \le i \le n$. Let for $1 \le a \le n/2$, $1 \le t \le n$,

$$\psi_{a,t}(i) = \frac{1}{\sqrt{2a}}(\chi_{[0,a]}(i - t + a) - \chi_{[a,2a]}(i - t + a)), \quad 1 \le i \le n, \tag{5}$$

where the expression $i - t + a$ is to be understood modulo n. We denote by

$$\mathcal{W} := \{\psi_{a,t}, 1 \le t \le n, 1 \le a \le n/2\}$$

the continuous Haar wavelet system. The cardinality of the continuous Haar wavelet system is $n^2/2$. We choose the continuous Haar wavelet system as a model for jumps in the signal i.e. a signal has a jump of size A at t, if $|\langle \xi, \psi_{a,t} \rangle| = A$ for some $1 \le a \le n/2$.

Of course our jump singularity model is not the only possibility. One could certainly also work with unbalanced jumps, which means that we only need a smooth part with small variation on one side of the jump. Should one take the continuous Haar wavelet transform of white noise, i.e. look at the values $\langle \psi_{a,t}, z \rangle$, a simple calculation yields that we obtain $n^2/2$ values that are normal distributed. If the values $\langle \psi_{a,t}, x \rangle$ where to be independent, the needle in a haystack problem suggest, that detection of a jump is only reliably possible if the jump size is larger than $\sqrt{2\log(n^2)}$. However, the probability variables $\langle \psi_{a,t}, x \rangle$ are not independent and we will show that the optimal threshold for reliable detection is $\sup_{a,t} |\langle \psi_{a,t}, x \rangle| > \sqrt{2\log(n)}$. This means, that if for some $\eta > 0$, $\sup_{a,t} |\langle \psi_{a,t}, x \rangle| \le (1 - \eta)\sqrt{2\log(n)}$ there does not exist an asymptotically powerful test, but if $\sup_{a,t} |\langle \psi_{a,t}, x \rangle| \ge (1 + \eta)\sqrt{2\log(n)}$ we can construct an asymptotically powerful test.

The key to show that these lower thresholds can be achieved is to subsample the scaling parameter of the continuous Haar wavelet system. For $\epsilon > 0$, we choose values $a_i = 2^{\epsilon k}, k = 1, \ldots \lceil \frac{1}{\epsilon}(\log(n) - 1) \rceil$ and define

$$\mathcal{W}_\epsilon := \left\{ \psi_{a_k,t}, k = 1, \ldots \lceil \frac{1}{\epsilon}(\log(n) - 1) \rceil, 1 \le t \le n \right\}.$$

We use the metric: $\delta(\xi_1, \xi_2) := (1 - \langle \xi_1, \xi_2 \rangle)^{\frac{1}{2}}$ to describe the difference between the elements of $\mathcal{W}_\epsilon$ and the continuous Haar wavelet system $\mathcal{W}$. The following Lemma shows, that the subsampled system is in some sense dense with respect to the metric δ.

Lemma 2.2. *For $\epsilon > 0$ the subsampled Haar wavelet system $\mathcal{W}_\epsilon$ is an ϵ-net for the continuous Haar wavelet system $\mathcal{W}$, i.e.*

$$\max_{a,t} \min_{k,s} \delta(\psi_{a_k,s}, \psi_{a,t}) < \epsilon.$$

3

Proof. For given $a \leq n/2$, there exists $a_k > 0$ such that either $a_k \leq a \leq 2^{\epsilon/2} a_k$ or $2^{\epsilon/2} a_k \leq a \leq a_{k+1}$. Then

$$\delta(\psi_{a,t}, \psi_{a_k,t})^2 = 1 - \langle \psi_{a,t}, \psi_{a_k,t} \rangle,$$

where

$$\langle \psi_{a,t}, \psi_{a_k,t} \rangle = \langle \psi_{a_k,t}, \psi_{a_k,t} \rangle + \langle \psi_{a_k,t} - \psi_{a,t}, \psi_{a_k,t} \rangle = 1 + \langle \psi_{a_k,t} - \psi_{a,t}, \psi_{a_k,t} \rangle.$$

Furthermore

$$\langle \psi_{a_k,t} - \psi_{a,t}, \psi_{a_k,t} \rangle \leq \|\psi_{a_k,t} - \psi_{a,t}\| \|\psi_{a_k,t}\| \leq \|\psi_{a_k,t} - \psi_{a,t}\|.$$

Assume $a_k \leq a \leq 2^{\epsilon/2} a_k$, then applying the definition (5) we see that $\|\psi_{a_k,t} - \psi_{a,t}\|^2$ can be estimated by $2\frac{1}{2a}|a - a_k| + a_k|\frac{1}{\sqrt{2a}} - \frac{1}{\sqrt{2a_k}}|^2$. Now we have

$$|a - a_k| \leq a_k(2^{\epsilon/2} - 1) < \epsilon/2.$$

This yields that

$$|1 - \langle \psi_{a,t}, \psi_{a_k,t} \rangle| \leq \epsilon^2$$

and hence $\delta(\psi_{a,t}, \psi_{a_k,t}) < \epsilon$.

The case of $2^{\epsilon/2} a_k \leq a \leq a_{k+1}$ follows analogously. $\square$

We denote by

$$\mathcal{W}(x) = \sup\{|\langle \psi_{a,t}, x \rangle|, \psi_{a,t} \in \mathcal{W})\}$$

the maximum value of the Haar coefficients of a signal x. Similarly

$$\mathcal{W}_\epsilon(x) = \sup\{|\langle \psi_{a,t}, x \rangle|, \psi_{a,t} \in \mathcal{W}_\epsilon)\}$$

denotes the maximum Haar coefficient of x with respect to the subsampled system $\mathcal{W}_\epsilon$. Using the ϵ-net of Lemma 2.2 we aim to now construct a reliable test. Indeed we will show, that if the jump size is larger than $\sqrt{2(1 + \eta)\log(n)}$ for some positive η the *generalized likelihood ratio test* (GLRT), which rejects the null hypothesis if

$$\mathcal{W}(x) \geq t_n,$$

is asymptotically powerful, if we choose $t_n/\sqrt{2\log(n)} \to 1$. To prove this, we employ the fact that the subsampled Haar wavelet system yields an ϵ-net. Furthermore, we make use of the *GLRT* with the reduced dictionary, i.e. we reject the null hypothesis if

$$\mathcal{W}_\epsilon(x) \geq t_n, \tag{6}$$

where $t_n/\sqrt{2\log(n)} \to 1$. The last ingredient for our desired result is the following result describing extreme values of Gaussian processes.

Lemma 2.3. *[2] Let $\omega_1, \ldots, \omega_m$ be (possibly dependent) $\mathcal{N}(0, \sigma_i^2)$ variables with all $\sigma_i \leq \sigma$. Then,*

$$P(\max\{\omega_1, \ldots, \omega_m\} > \sqrt{2\log(m)}\sigma \leq \frac{1}{\sqrt{4\pi \log(m)}}.$$

4

Now we will prove that asymptotically the reduced GLRT and the full GLRT behave in the same way. For this, let

$$\delta_\epsilon(x) := \min_{\psi_{a_k,t} \in \mathcal{W}_\epsilon} |\mathcal{W}(x) - \langle \psi_{a_k,t}, x \rangle|,$$

denote the effective distance of $\mathcal{W}$ to $\mathcal{W}_\epsilon$.

Lemma 2.4. *For each $\eta > 0$,*

$$P_{H_{0,n}} \left\{ \delta_\epsilon(x) > \sqrt{8 \log(n)}\epsilon \right\} \leq \frac{1}{\sqrt{\log(n)}}.$$

Proof. First of all we assume that the maximum of the full GLRT is attained at $\psi_{a,t}$. By Lemma 2.2 we have that there exists $\psi_{a_k,\tilde{t}} \in \mathcal{W}_\epsilon$ such that $\|\psi_{a,t} - \psi_{a_k,\tilde{t}}\|_2 \leq \epsilon$. Then

$$\delta(x) \leq |\langle \psi_{a_k,\tilde{t}} - \psi_{a,t}, x \rangle| = |\sum_{1 \leq i \leq n} (\psi_{a_k,\tilde{t}}(i) - \psi_{a,t}(i))x(i)|.$$

Since we assume that $H_{0,n}$ holds, we have that $x(i) = z(i) \sim N(0,1)$ for all $1 \leq i \leq n$. Now $\sum_i (\psi_{a_k,\tilde{t}}(i) - \psi_{a,t}(i))z(i) \sim \mathcal{N}(0, \|\psi_{a_k,\tilde{t}}(i) - \psi_{a,t}\|^2) = \mathcal{N}(0, \tilde{\epsilon}^2)$ with $\tilde{\epsilon} \leq \epsilon$. To estimate the probability of $\delta_\epsilon(x)$ exceeding $4\sqrt{\log(n)}\epsilon$ we have to take the maximum of all probability variables $\sum_i (\psi_{a_k,\tilde{t}}(i) - \psi_{a,t}(i))z(i)$ over all a, t and corresponding $a_k, \tilde{t}$. These are certainly less than n^4 variables which are all normal distributed with variance less than ϵ^2. Hence Lemma 2.3 implies that

$$P_{H_{0,n}} \left\{ \delta_\epsilon(x) > \sqrt{2 \log(n^4)}\epsilon \right\} \leq \frac{1}{\sqrt{4\pi \log(n^4)}} \leq \frac{1}{\sqrt{\log(n)}}.$$

$\square$

Now we are in the position to prove that the GLRT is asymptotically powerful, should the jump size exceed $\sqrt{2(1 + \eta) \log(n)}$.

Theorem 2.5. *For each $\eta > 0$,*

$$P_{H_0} \left\{ \mathcal{W}(x) > \sqrt{2(1 + \eta) \log(n)} \right\} \to 0, \quad for \ n \to \infty.$$

Proof. Let us first prove that, for $\epsilon > 0$,

$$P_{H_0} \left\{ \mathcal{W}_\epsilon(x) > \sqrt{2(1 + \eta) \log(n)} \right\} \to 0, \quad for \ n \to \infty.$$

In fact $\#\mathcal{W}_\epsilon \leq \frac{1}{\epsilon} n(\log(n) - 1)$. Since under H_0, $\langle \psi_{a_k,t}, x \rangle \sim \mathcal{N}(0,1)$ for all $\psi_{a_k,t} \in \mathcal{W}_\epsilon$ Lemma 2.3 implies that

$$P_{H_0} \left\{ \mathcal{W}_\epsilon(x) > \sqrt{2 \log(\frac{1}{\epsilon} n(\log(n) - 1))} \right\} = P_{H_0} \left\{ \mathcal{W}_\epsilon(x) > \sqrt{2(\log(n) + \log(\log(n) - 1) - \log(\epsilon))} \right\}$$

$$\leq \frac{1}{\sqrt{4\pi \log(\frac{1}{\epsilon} n(\log(n) - 1))}}.$$

5

Furthermore we obtain that for any $\eta > 0$, there exists $M_\eta > 0$ such that for all $m \geq M_\eta$ we have $\eta \log(m) \geq \log(\log(m) - 1) - \log(\epsilon)$. Hence

$$P_{H_0}\left\{\mathcal{W}_\epsilon(x) > \sqrt{2(1+\eta)\log(m)}\right\} \leq P_{H_0}\left\{\mathcal{W}_\epsilon(x) > \sqrt{2(\log(n) + \log(\log(n) - 1) - \log(\epsilon))}\right\}$$

$$\leq \frac{1}{\sqrt{4\pi \log(n)}} \to 0, \quad \text{for } n \to \infty.$$

The result for $\mathcal{W}(x)$ follows by choosing ϵ such that $\sqrt{2(1+\eta)} - \sqrt{8}\epsilon \leq \sqrt{2(1+\eta/2)}$, then Lemma 2.4 and the observation that

$$\mathcal{W}_\epsilon(x) \leq \mathcal{W}(x) \leq \mathcal{W}_\epsilon(x) + \delta_\epsilon(\xi),$$

imply the result. □

Of course the question arises weather the bound of Theorem 2.5 is tight, i.e. if $\sqrt{2(1+\eta)\log(n)}$ is the minimum jump size that can be detected. The following Lemma states that this is indeed the case, i.e. that a lower jump size will not be detectable with an asymptotically powerful test.

Lemma 2.6. *Given a sequence of hypotheses*

$$H_{1,n}^\eta : x(i) = \xi_J(i) + z(i), \quad 1 \leq i \leq n,$$

and $\eta > 0$ with

$$\mathcal{W}_\epsilon(\xi_J) = \sqrt{2(1-\eta)\log(n)}.$$

Then, no test is asymptotically powerful for testing $H_{0,n}$ against $H_{1,n}^\eta$.

Proof. Let $H_{0,n}$ be the hypothesis that $x(i) \sim \mathcal{N}(0, 1)$. Furthermore let $H_{1,n}'$ be the hypothesis that for any $0 \leq t \leq n/2 - 1$

$$x(i) = A_n \psi_{1,2t} + z(i), z_i \sim \mathcal{N}(0, 1), \quad 0 \leq i \leq n - 1.$$

Consider the orthonormal Haar wavelet basis on $1 \leq i \leq n$ consisting of

$$\left\{\psi_{1,2t} : t \in [0, n/2]\right\} \cup \left\{\frac{1}{\sqrt{2}}\chi_{[2t,2t+1]} : t \in [0, n/2]\right\}.$$

Indeed, since this yields an orthonormal basis for the signals on I_n we can apply a basis transformation and represent x with respect to the Haar wavelet basis to obtain $\hat{x}$. Now $H_{1,n}'$ becomes

$$\hat{x}(i) = A_n \delta_t(i) + z(i), \quad 0 \leq i \leq n.$$

The needle in a haystack problem, suggests, that if $A_n = \sqrt{2(1 - \eta \log(n))}$ for $\eta > 0$, then $H_{0,n}$ and $H_{1,n}'$ merge asymptotically, i.e. there does not exist an asymptotically powerful test to distinguish between the two hypotheses. Since distinguishing $H_{1,n}'$ from $H_{0,n}$ should be easier than distinguishing $H_{1,n}$ from $H_{0,n}$ this proves the claim. □

3 General Geometrical Structures

3.1 General Framework

In this section we will extract the main ideas of the path we followed in the previous subsection on Haar wavelets to give a theory for arbitrary systems of functions. This can then be employed in any dimension and especially in Section 3.2 to two dimensional functions. We saw in the Haar wavelet case, that subsampling the functions dyadically does not destroy the detectability of jumps by that system. Indeed, we will show, that whenever we have set of functions on I_n: $(g_j)_{j \in J_n}$ with some index set J_n such that for every $\epsilon > 0$ there exist subsystems, that yield an ϵ-net the reduced GLRT is asymptotically powerful if the signal has a *detectable feature*:

Definition 3.1. *Given a function system $(g_j)_{j \in J_n}$ with a subsystem $(g_j)_{j \in J_n^\epsilon}, J_n^\epsilon \subset J_n$ on I_n. If for each $l \in J_n$ we have that $\min_{j \in J_n^\epsilon} \|g_j - g_l\| < \epsilon$, we say that $(g_j)_{j \in J_n^\epsilon}$ constitutes an ϵ-net for $(g_j)_{j \in J_n}$.*

Definition 3.2. *For each $n \in \mathbb{N}$ let $(g_j)_{j \in J_n}$ be a system of functions on I_n with $\|g_j\|_{\ell^2} = \sum_{i \in I_n} |g_j(i)|^2 = 1$. Furthermore assume that for every ϵ there exists a subsystem $((g_j)_{j \in J_n^\epsilon})$ hat constitutes an ϵ-net for $(g_j)_{j \in J_n}$. Assume that there exists $e > 0$ such that for every $\omega > 0$ there exists a constant C_ω with*

$$|J_n^\epsilon| \leq C_\omega n^{e+\omega}, \quad \text{for all } n \in N.$$

A sequence of signals $\xi^{(n)}$ has a detectable feature *if $\sup_{j \in J_n} |\langle \xi^{(n)}, g_j \rangle| \geq (1+\eta) \sqrt{2 \log(n^e)}$ for some $\eta > 0$ and $n \in \mathbb{N}$.*

The jump singularities from the preceding subsection are detectable features of the continuous Haar wavelet system $\mathcal{W}$. We will study another examples in the sequel which are detectable features of continuous shearlet systems. Moving on with the general set-up and a set of functions $(g_j)_{j \in J_n}$, we define

$$\mathcal{G}_n(x) := \max_{j \in J_n} |\langle g_j, x \rangle|,$$

and for the subsystems we write

$$\mathcal{G}_n^\epsilon(x) := \max_{j \in J_n^\epsilon} |\langle g_j, x \rangle|.$$

$\mathcal{G}_n(x)$ and $\mathcal{G}_n^\epsilon(x)$ denote the maximum response from a signal, when tested with $\mathcal{G}_n$ or $\mathcal{G}_n^\epsilon$ respectively. Using these we can again build a full and a reduced GLRT by checking, whether $\mathcal{G}_n(x)$ or $\mathcal{G}_n^\epsilon(x)$ exceed a critical threshold. Furthermore, we aim to obtain an analog of Lemma 2.4, which would tell us that asymptotically we can replace the full GLRT with the reduced GLRT. With

$$\delta_\epsilon(x) := \min_{j \in J_n^\epsilon} |\mathcal{G}(x) - \langle g_j, x \rangle|,$$

we obtain

Lemma 3.3. *For each $n \in \mathbb{N}$ let $(g_j)_{j \in J_n}$ be a set of functions on I_n. Assume that for every $\epsilon > 0$ we have subsystems $((g_j)_{j \in J_n^\epsilon})$ that yield an ϵ-net for $(g_j)_{j \in J_n}$. Then, we have*

$$P_{H_{0,n}} \left\{ \delta_\epsilon(x) > \sqrt{8 \log(n)} \epsilon \right\} \leq \frac{1}{\sqrt{\log(n)}}.$$

Proof. The proof is the same as that of Lemma 2.4. $\qquad\square$

Now we can show, that every detectable feature can, as the name already predicts, be detected. Furthermore it is possible to construct a test using only a reduced GLRT.

Theorem 3.4. *For every $n \in \mathbb{N}$ let $(g_j)_{j \in J_n}$ be a set of normalized functions with subsystems as in Definition 3.2 that have cardinality $O(n^{e+\omega})$ for $e > 0$ and every $\omega > 0$. Let $\xi^{(n)}$ be a sequence of signals that have a detectable feature and*

$$x^n(i) = \xi^{(n)}(i) + z(i), \ \text{where} \ z(i) \sim \mathcal{N}(0,1) \ \text{and} \ i \in I_n.$$

Then there exists $\epsilon > 0$ such that for all $\eta > 0$

$$P\left\{\mathcal{G}^\epsilon(z) > (\sqrt{2(1+\eta)\log(n^e)})\right\} \to 0, \quad \text{for } n \to \infty$$

and there exists $\eta > 0$ such that

$$P\left\{\mathcal{G}^\epsilon(x^{(n)}) > (\sqrt{2(1+\eta)\log(n^e)})\right\} \to 1, \quad \text{for } n \to \infty.$$

Proof. The first part of the proof uses similar arguments as the proof of Theorem 2.5 For $\epsilon, \eta > 0$, there exists $\omega < \eta$ and C_ω such that $\#\mathcal{G}_\epsilon \leq C_\omega n^{e+\omega}$. Since the g_j are normalized we obtain that $\langle g_j, z \rangle \sim \mathcal{N}(0,1)$. Lemma 2.3 implies

$$P\left\{\mathcal{W}_\epsilon(z) > \sqrt{2\log(C_\omega n^{e+\omega})}\right\} = P\left\{\mathcal{W}_\epsilon(z) > \sqrt{2(1+\omega)\log(n^e) - \log(C_\omega)}\right\}$$
$$\leq \frac{1}{\sqrt{4\pi \log(C_\omega n^{e+\omega})}}.$$

Furthermore we obtain that there exists $M_\eta > 0$ such that for $n \geq M_\eta$ we have $\eta \log(n^e) \geq \omega \log(n^e) - \log(C_\omega)$. Hence

$$P\left\{\mathcal{W}_\epsilon(z) > \sqrt{2(1+\eta)\log(n^e)}\right\} \leq P\left\{\mathcal{W}_\epsilon(z) > \sqrt{2(1+\omega)\log(n^e) - \log(C_\omega)}\right\}$$
$$\leq \frac{1}{\sqrt{4\pi \log(n^e)}} \to 0.$$

For the second part we first prove that there exists some $\eta > 0$ such that

$$P\left\{\mathcal{G}(x^{(n)}) > (\sqrt{2(1+\eta)\log(n)})\right\} \to 1, \quad \text{for } n \to \infty. \tag{7}$$

By the assumption that x has a detectable feature we obtain that there exists some γ such that $\mathcal{G}(x^{(n)}) \sim \mathcal{N}(\sqrt{2(1+\gamma)\log(n^e)}, 1)$. Choosing $\eta < \gamma$ less than γ yields that (7) is equal to

$$P\left\{\mathcal{N}(\sqrt{2(1+\gamma)\log(n^e)}, 1) > \sqrt{2(1+\eta)\log(n)}\right\} = P\left\{\mathcal{N}(0,1) > -\beta\sqrt{\log n}\right\} \leq n^{-\beta^2/2} \to 0,$$

where the last equation follows by Mills ratio and $\beta = \sqrt{2(1+\eta)} - \sqrt{2(1+\gamma)}$. To obtain that

$$P\left\{\mathcal{G}^\epsilon(x^{(n)}) > (\sqrt{2(1+\eta)\log(n)})\right\}$$

8

we observe that

$$\delta_\epsilon(x^{(n)}) = \min_{j \in J_n^\epsilon} |\mathcal{G}(x^{(n)}) - \left\langle g_j, x^{(n)} \right\rangle| = \min_{j \in J_n^\epsilon} |\mathcal{G}(\xi^{(n)}) - \left\langle g_j, \xi^{(n)} \right\rangle + \mathcal{G}(z) - \left\langle g_j, z \right\rangle|$$
$$\leq \min_{j \in J_n^\epsilon} (|\mathcal{G}(\xi^{(n)}) - \left\langle g_j, \xi^{(n)} \right\rangle| + |\mathcal{G}(z) - \left\langle g_j, z \right\rangle|).$$

Let $g_{\tilde{j}} \in \mathcal{G}$ be a function where $|\langle g_{\tilde{j}}, \xi^{(n)} \rangle| = \mathcal{G}(\xi^{(n)})$, then

$$\min_{j \in J_n^\epsilon} (|\mathcal{G}(\xi^{(n)}) - \left\langle g_j, \xi^{(n)} \right\rangle| + |\mathcal{G}(z) - \left\langle g_j, z \right\rangle|) \leq |\left\langle g_{\tilde{j}} - g_j, \xi^{(n)} \right\rangle| + |\left\langle g_{\tilde{j}} - g_j, z \right\rangle|$$
$$\leq \epsilon + |\left\langle g_{\tilde{j}} - g_j, z \right\rangle|.$$

The rest follows by Lemma 3.3. $\qquad\square$

Let us reflect shortly on what we proved so far. If a signal has a detectable structure, i.e., a structure that can be detected by a set of functions, we can subsample this set to obtain a cheaper test. We saw analyzed how this can be applied for Haar wavelets. Jump singularities are detectable by the continuous Haar wavelet system as we observed in Subsection 2.2. Furthermore the continuous Haar wavelet system can be subsampled to yield ϵ-nets as shown in Lemma 2.2. This smaller set allows for the construction of a fast test by computing all the scalar products of the subsampled Haar wavelet coefficients with the signal. In fact this can be done in $O(n \log(n))$, which we see by observing, that $\langle \psi_{a_k,t}, x \rangle = \psi_{a,0} * x(t) = \text{ifft}(\text{fft}(\psi_{a,0})\text{fft}(x))$, where fft and ifft are the discrete fast Fourier transform and its inverse. Since we have only $O(\log(n))$ scaling parameters, this yields the claimed computational complexity of $O(n \log(n))$. Since this machinery seems quite powerful we can now advance to higher dimensional structures namely those described by continuous shearlet systems.

3.2 Detecting 2D Jump Singularities

From the work of [9] is is clear, that continuous shearlet systems are well versed to detect geometric structures like discontinuities along smooth curves. By virtue of Theorem 3.4 we need to show, that jump singularities are a detectable feature of a continuous shearlet system and that subsampling the shearlet system yields a subsystem, that constitutes an ϵ-net.

Let us start by defining the continuous digital shearlet transform on a $n \times n$ pixel image. As the digital realm should be a discretization of the continuum realm $L^2([0,1]^2)$ this should be reflected in the construction. Let ψ be a continuous shearlet supported in $[0,1]^2$. We define the continuous digital shearlet system as digitization of the continuous system by defining for an $n \times n$ pixel image the continuous digital shearlet system $\mathcal{CDSH}$ as

$$\mathcal{CDSH} := \left\{ \psi_{a,s,t}, a = 1, \dots n, s = -\frac{n}{2}, \dots, \frac{n}{2}, t = (t_1, t_2), 1 \leq t_i \leq n \right\},$$

where

$$\psi_{a,s,t}(i) = n^{\frac{5}{4}} \int_{x \in Q_{\frac{1}{2n}}(i/n)} a^{-\frac{3}{4}} \psi(A_{\frac{a}{n}}^{-1} S_{\frac{s}{n}}^{-1}(x - t/n))dx.$$

In the equation above $Q_q(r)$ denotes the cube of sidelength q and center r. As a second step we define subsystems with significantly lower complexity that still yield ϵ−nets. With $\lceil t \rceil = (\lceil t_1 \rceil, \lceil t_2 \rceil)$, we define for $\eta, \omega, \nu > 0$,

$$\mathcal{CDSH}^{\delta,\omega,\nu} := \left\{ \psi_{\lceil a_k \rceil, \lceil s \rceil, \lceil t \rceil}, \ a_k = 2^{\delta k}, \ s = \omega\sqrt{a_k n}l, \ t = (t_1, t_2), \ t_1 = \nu a_k r_1, \ t_2 = \nu\sqrt{a_k n}r_2, \right.$$

$$k = 0, \ldots, \left\lceil \frac{1}{\delta}(\log(n) - 1) \right\rceil, \ l = -\left\lfloor \frac{\sqrt{n}}{\sqrt{a_k}\omega} \right\rfloor, \ldots, \left\lfloor \frac{\sqrt{n}}{\sqrt{a_k}\omega} \right\rfloor,$$

$$\left. r_1 = 0, \ldots, \frac{n}{a_k \nu}, \ , \ r_2 = 0, \ldots, \frac{\sqrt{n}}{\sqrt{a_k}\nu} \right\}.$$

Observe that the cardinality of $\mathcal{CDSH}^{\delta,\omega,\nu}$ is of order $n^2(\log n)$, which compared to the colossal $O(n^4)$ of the $\mathcal{CDSH}$ can be seen as a significant reduction. In order to apply Theorem 3.4 we still need to see that the new system constitutes an ϵ-net.

Lemma 3.5. *Let ψ be a Lipschitz continuous shearlet with Lipschitz constant C_ψ. Then for $\delta \leq 1$, $\mathcal{CDSH}^{\delta,\omega,\nu}$ is an ϵ-net for $\mathcal{CDSH}$ with $\epsilon = 2\delta/3 + C_\psi 3\delta + \omega + 3C_\psi\nu$.*

Proof. We compute

$$\|\psi_{a,s,t} - \psi_{a_k,\tilde{s},\tilde{t}}\|,$$

where we need to choose k, $\tilde{s}$ and $\tilde{t}$ in a suitable way. Without loss of generality, we can assume $t = 0$, since the above expression is invariant under translations. Furthermore, we choose

$$a_k = \arg\min_k |\lceil a_k \rceil - a|, \quad \tilde{s} = \arg\min_{\hat{s} = -\frac{n}{2}: \omega\sqrt{a_k n}: \frac{n}{2}} |s - \lceil \hat{s} \rceil|.$$

Assume for simplicity that $a_k < a$. Then we obtain the error estimates

$$|\lceil a_k \rceil - a| \leq a_k(2^\delta - 1), \quad |s - \lceil \tilde{s} \rceil| \leq \omega\sqrt{a_k n}/2.$$

Moreover, choose $\tilde{t} = (\tilde{t}_1, \tilde{t}_2)$ such that

$$|\tilde{t}_1| \leq \frac{\nu}{2}a_k, \quad |\tilde{t}_2| \leq \frac{\nu}{2}\sqrt{a_k n}.$$

Then we obtain

$$\|\psi_{a,s,0} - \psi_{a_k,\tilde{s},\tilde{t}}\|^2$$
$$= \sum_{i \in I_n} \left| \psi_{a,s,0}(i) - \psi_{a_k,\tilde{s},\tilde{t}}(i) \right|^2$$
$$= \sum_{i \in I_n} \left| n^{\frac{5}{4}} \int_{x \in Q_{\frac{1}{2n}}(i/n)} a^{-\frac{3}{4}}\psi(A_{\frac{a}{n}}^{-1}S_{\frac{s}{n}}^{-1}x) - \lceil a_k \rceil^{\frac{3}{4}}\psi(A_{\frac{\lceil a_k \rceil}{n}}^{-1}S_{\frac{\lceil \tilde{s} \rceil}{n}}^{-1}(x - \tilde{t}/n))dx \right|^2$$
$$= \sum_{i \in I_n} \left| n^{\frac{5}{4}} \int_{x \in Q_{\frac{1}{2n}}(i/n)} a^{-\frac{3}{4}}\psi(A_{\frac{a}{n}}^{-1}S_{\frac{s}{n}}^{-1}x) - \lceil a_k \rceil^{-\frac{3}{4}}\psi(A_{\frac{\lceil a_k \rceil}{n}}^{-1}S_{\frac{\lceil \tilde{s} \rceil}{n}}^{-1}x - A_{\frac{\lceil a_k \rceil}{n}}^{-1}S_{\frac{\lceil \tilde{s} \rceil}{n}}^{-1}\tilde{t}/n)dx \right|^2.$$

10

Now we apply the transformation $x \mapsto S_{\frac{\lceil \tilde{s} \rceil}{n}} A_{\frac{\lceil a_k \rceil}{n}} x$ to the last term above. We obtain that

$$\|\psi_{a,s,0} - \psi_{a_k,\tilde{s},\tilde{t}}\|^2 = \sum_{i \in I_n} |n^{-\frac{1}{2}} \int_{D_i} \psi(x) - \left(\frac{a}{\lceil a_k \rceil}\right)^{\frac{3}{4}} \psi(A^{-1}_{\frac{\lceil a_k \rceil}{n}} S_{\frac{s-\lceil \tilde{s} \rceil}{n}} A_{\frac{a}{n}} x - A^{-1}_{\frac{\lceil a_k \rceil}{n}} S^{-1}_{\frac{\lceil \tilde{s} \rceil}{n}} \tilde{t}/n) dx|^2,$$

where $D_i = \left(S_{\frac{\lceil \tilde{s} \rceil}{n}} A_{\frac{\lceil a_k \rceil}{n}}\right)^{-1} Q_{\frac{1}{2n}}(i/n)$. Since $|D_i| \leq \frac{1}{\sqrt{n}}$ we only need to show that for $x \in D_i$

$$|\psi(x) - \left(\frac{a}{\lceil a_k \rceil}\right)^{\frac{3}{4}} \psi(A^{-1}_{\frac{\lceil a_k \rceil}{n}} S_{\frac{s-\lceil \tilde{s} \rceil}{n}} A_{\frac{a}{n}} x - A^{-1}_{\frac{\lceil a_k \rceil}{n}} S^{-1}_{\frac{\lceil \tilde{s} \rceil}{n}} \tilde{t}/n)|^2 < 2\delta/3 + C_\psi 3\delta + \omega + 3C_\psi \nu.$$

We estimate as follows:

$$\left| \psi(x) - \left(\frac{a}{\lceil a_k \rceil}\right)^{\frac{3}{4}} \psi(A^{-1}_{\frac{\lceil a_k \rceil}{n}} S_{\frac{s-\lceil \tilde{s} \rceil}{n}} A_{\frac{a}{n}} x - A^{-1}_{\frac{\lceil a_k \rceil}{n}} S^{-1}_{\frac{\lceil \tilde{s} \rceil}{n}} \tilde{t}/n) \right| \leq I_1 + I_2 + I_3,$$

where

$$I_1 := \left| \psi(x) - \left(\frac{a}{\lceil a_k \rceil}\right)^{\frac{3}{4}} \psi(x) \right|$$

$$I_2 := \left(\frac{a}{\lceil a_k \rceil}\right)^{\frac{3}{4}} \left| \psi(x) - \psi(A^{-1}_{\frac{\lceil a_k \rceil}{n}} S_{\frac{s-\lceil \tilde{s} \rceil}{n}} A_{\frac{a}{n}} x) \right|$$

$$I_3 := \left(\frac{a}{\lceil a_k \rceil}\right)^{\frac{3}{4}} \left| \psi(A^{-1}_{\frac{\lceil a_k \rceil}{n}} S_{\frac{s-\lceil \tilde{s} \rceil}{n}} A_{\frac{a}{n}} x) - \psi(A^{-1}_{\frac{\lceil a_k \rceil}{n}} S_{\frac{s-\lceil \tilde{s} \rceil}{n}} A_{\frac{a}{n}} x - A^{-1}_{\frac{\lceil a_k \rceil}{n}} S^{-1}_{\frac{\lceil \tilde{s} \rceil}{n}} \tilde{t}/n) \right|.$$

Clearly $I_1 \leq (2^{\frac{3}{4}\delta} - 1) \leq 2\delta/3$. The cases I_2, I_3 follow by the Lipschitz continuity of ψ by

$$I_2 \leq 2^{\frac{3}{4}\delta} C_\psi |x - A^{-1}_{\frac{\lceil a_k \rceil}{n}} S_{\frac{s-\lceil \tilde{s} \rceil}{n}} A_{\frac{a}{n}} x| \text{ and } I_3 \qquad \leq 2^{\frac{3}{4}\delta} C_\psi |A^{-1}_{\frac{\lceil a_k \rceil}{n}} S^{-1}_{\frac{\lceil \tilde{s} \rceil}{n}} \tilde{t}/n|.$$

First of all

$$A^{-1}_{\frac{\lceil a_k \rceil}{n}} S_{\frac{s-\lceil \tilde{s} \rceil}{n}} A_{\frac{a}{n}} = \begin{pmatrix} \frac{n}{\lceil a_k \rceil} & 0 \\ 0 & \sqrt{\frac{n}{\lceil a_k \rceil}} \end{pmatrix} \begin{pmatrix} 1 & \frac{s-\lceil \tilde{s} \rceil}{n} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{a}{n} & 0 \\ 0 & \sqrt{\frac{a}{n}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{n}{\lceil a_k \rceil} & \frac{s-\lceil \tilde{s} \rceil}{\lceil a_k \rceil} \\ 0 & \sqrt{\frac{n}{\lceil a_k \rceil}} \end{pmatrix} \begin{pmatrix} \frac{a}{n} & 0 \\ 0 & \sqrt{\frac{a}{n}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{a}{\lceil a_k \rceil} & \frac{(s-\lceil \tilde{s} \rceil)\sqrt{a}}{\lceil a_k \rceil \sqrt{n}} \\ 0 & \sqrt{\frac{a}{\lceil a_k \rceil}} \end{pmatrix}.$$

This implies that $I_2 \leq 2^{\frac{3}{4}\delta} C_\psi (|x - \frac{a}{\lceil a_k \rceil} x| + \frac{(s-\lceil \tilde{s} \rceil)\sqrt{a}}{\lceil a_k \rceil \sqrt{n}} |x_2|)$. Using that $\frac{a}{\lceil a_k \rceil} = 2^\delta$ and $\frac{(s-\lceil \tilde{s} \rceil)\sqrt{a}}{\lceil a_k \rceil \sqrt{n}} \leq \omega/2$ we obtain that $I_2 \leq C_\psi 2^{\frac{3}{4}\delta}((2^\delta - 1) + \omega/2)\|x\|_2$ and since $\|x\|_2 \leq \sqrt{2}$ we get $I_2 \leq C_\psi 2^{\frac{3}{4}\delta + \frac{1}{2}}((2^\delta - 1) + \omega/2) \leq C_\psi 3\delta + \omega$.

11

With a similar approach we can estimate I_3 by

$$2^{\frac{3}{4}\delta} C_\psi \begin{pmatrix} \frac{n}{\lceil a_k \rceil} & -\frac{\lceil \tilde{s} \rceil}{\lceil a_k \rceil} \\ 0 & \sqrt{\frac{n}{\lceil a_k \rceil}} \end{pmatrix} \tilde{t}/n = 2^{\frac{3}{4}\delta} C_\psi \begin{pmatrix} \frac{1}{\lceil a_k \rceil} & -\frac{\lceil \tilde{s} \rceil}{n\lceil a_k \rceil} \\ 0 & \frac{1}{\sqrt{n\lceil a_k \rceil}} \end{pmatrix} \tilde{t}$$

$$\leq 2^{\frac{3}{4}\delta} C_\psi \left(\frac{a_k}{\lceil a_k \rceil} \frac{\nu}{2} + \frac{\sqrt{a_k}\lceil \tilde{s} \rceil}{\sqrt{n}\lceil a_k \rceil} \frac{\nu}{2} + \frac{\sqrt{a_k}}{\sqrt{\lceil a_k \rceil}} \frac{\nu}{2} \right) \leq 3 C_\psi \nu.$$

In conclusion we obtain that

$$I_1 + I_2 + I_3 \leq 2\delta/3 + C_\psi 3\delta + \omega + 3C_\psi \nu$$

which yields the claim. $\qquad\square$

Next we can combine Theorem 3.4 and Lemma 3.5 to obtain, that subsampled shearlet systems can detect all detectable features of unsubsampled shearlet systems.

Theorem 3.6. *Let ψ be a Lipschitz continuous shearlet and $\mathcal{CDSH}$ the corresponding continuous digital shearlet system. Let $\xi^{(n)}$ be a sequence of signals that have a detectable feature. Then, there exists $\delta, \omega, \nu > 0$ such that the GLRT with $\mathcal{CDSH}^{\delta,\omega,\nu}$ is asymptotically powerful.*

3.3 Open Questions

What can be detected by $\mathcal{CDSH}$?
First of all it is not yet clear, what type of geometrical structures can be detected with the $\mathcal{CDSH}$. Certainly, elements of $\mathcal{CDSH}$ are detectable with the GLRT and a reduced shearlet system $\mathcal{CDSH}^{\delta,\omega,\nu}$, but one would certainly be interested in objects or structures that appear naturally in images. Since the publications [3, 4, 9] show how shearlets can detect jump singularities, an extension to these structures seems very likely.

Is the subsampling optimal?
We see that the cardinality of the subsampled shearlet systems $\mathcal{CDSH}^{\delta,\omega,\nu}$ is $O(n^2 \log(n))$. Given that the image has n^2 pixels, it seems very likely that, at least up to the log factor, this is the minimal cardinality for which an ϵ-net can be constructed. A proof for this optimality is still missing. A different approach would be to allow further subsampling on the cost of an increasing signal strength. If this is possible, a precise characterization of the trade-off as in [5] is desirable.

Extensions
Similar to the question of what can be detected, we can also adjust the process of testing. For instance if we assume we have an image composed out of only a few elements of $\mathcal{CDSH}$, we can develop a procedure, where we find some promising elements in $\mathcal{CDSH}^{\delta,\omega,\nu}$ and try various combinations of them. A similar method to extract curves from noisy images based on extensions of promising parts of curves has been proposed in [1]. Since shearlets yield optimally sparse approximations of cartoon-like images [8], this might ultimately give rise to a method to optimally detect cartoon-like images embedded in noise.

References

[1] S. Alpert, M. Galun, B. Nadler, and R. Basri. Detecting faint curved edges in noisy images. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision – ECCV 2010*, volume 6314 of *Lecture Notes in Computer Science*, pages 750–763. Springer Berlin Heidelberg, 2010.

[2] E. Arias-Castro, D. L. Donoho, and X. Huo. Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory*, 51(7):2402–2425, 2005.

[3] K. Guo and D. Labate. Characterization and analysis of edges using the continuous shearlet transform. *SIAM J. Imaging Sci.*, 2(3):959–986, 2009.

[4] K. Guo, D. Labate, and W.-Q. Lim. Edge analysis and identification using the continuous shearlet transform. *Appl. Comput. Harmon. Anal.*, 27(1):24–46, 2009.

[5] I. Horev, B. Nadler, E. Arias-Castro, M. Galun, and R. Basri. Edge detection on a computational budget: A sublinear approach. *SIAM J. Imaging Sci.*, 8(1):458 –483.

[6] I. Ibragimov and R. Khas'minskii. *Statistical Estimation: Asymptotic Theory*. Springer, New York.

[7] I. A. Ibragimov, A. S. Nemirovskii, and R. Z. Khas'minskiii. Some problems of nonparametric estimation in gaussian white noise. *Theory of Probability and its Applications*, 31(3):391–406, 1986.

[8] G. Kutyniok and W.-Q. Lim. Compactly supported shearlets are optimally sparse. *J. Approx. Theory*, 163(11):1564–1589, 2011.

[9] G. Kutyniok and P. Petersen. Classification of edges using compactly supported shearlets. *Preprint*, 2014.